

TUGAS WEEK 3 – CASE ANALYSIS

PENDAHULUAN

1.1 Latar Belakang

Kanker paru merupakan penyebab utama kematian akibat kanker di seluruh dunia. Merokok bertanggung jawab terhadap sekitar 90% kasus kanker paru dan berkontribusi signifikan terhadap penurunan angka kelangsungan hidup pasien (Landi dkk. 2008). Meskipun hubungan antara merokok dan kanker paru telah lama diketahui, perubahan molekuler spesifik yang diinduksi oleh paparan rokok pada jaringan paru dan perannya dalam perkembangan serta prognosis kanker masih belum sepenuhnya dipahami. Perkembangan teknologi microarray memungkinkan analisis ekspresi gen secara genome-wide untuk mengidentifikasi pola ekspresi gen (gene expression signature) yang berhubungan dengan paparan lingkungan tertentu. Pendekatan ini memungkinkan identifikasi gen-gen yang mengalami perubahan regulasi akibat merokok serta penentuan jalur biologis yang terlibat dalam karsinogenesis paru.

Pada penelitian Landi dkk. (2008), analisis ekspresi gen dilakukan menggunakan platform Affymetrix HG-U133A pada sampel jaringan adenokarsinoma paru dan jaringan paru non-tumor berpasangan dari individu perokok aktif, mantan perokok, dan bukan perokok. Pendekatan statistik yang ketat digunakan untuk mengidentifikasi gen yang berbeda secara signifikan antar kelompok, termasuk penyesuaian terhadap faktor perancu, koreksi multiple testing menggunakan metode Benjamini-Hochberg untuk mengontrol False Discovery Rate (FDR), serta analisis Gene Ontology (GO) dan Gene Set Enrichment Analysis (GSEA) untuk klasifikasi fungsional. Melalui pendekatan ini, penelitian bertujuan mengidentifikasi signature ekspresi gen yang khas akibat merokok pada adenokarsinoma paru, mengevaluasi apakah perubahan tersebut juga terdapat pada jaringan paru non-tumor, serta menilai relevansinya terhadap perkembangan kanker dan kelangsungan hidup pasien.

1.2 Rumusan Masalah

1. Gen apa saja yang mengalami perubahan ekspresi signifikan (upregulation dan downregulation) antara kelompok yang dibandingkan?
2. Bagaimana pola ekspresi 50 gen dengan perubahan paling signifikan ketika divisualisasikan dalam bentuk heatmap?
3. Jalur biologis dan proses molekuler apa yang secara signifikan diperkaya (enriched) berdasarkan hasil analisis GO dan KEGG pathway?

1.3 Tujuan Analisis

Tujuan dari analisis ini adalah:

1. Mengidentifikasi Differentially Expressed Genes (DEGs) antara dua kelompok kondisi menggunakan pendekatan statistik berbasis limma.
2. Mengklasifikasikan gen-gen yang mengalami upregulation dan downregulation serta menampilkannya dalam bentuk volcano plot.
3. Menampilkan 50 DEGs teratas dalam bentuk heatmap untuk melihat pola ekspresi antar sampel.

4. Melakukan analisis enrichment menggunakan Gene Ontology (GO) dan KEGG pathway untuk mengidentifikasi jalur biologis yang terdampak secara signifikan.
5. Menginterpretasikan hasil analisis dalam konteks mekanisme biologis yang relevan dengan kondisi yang diteliti.

METODE

2.1 Dataset

Dataset yang digunakan dalam analisis ini diperoleh dari Gene Expression Omnibus (GEO) dan merupakan bagian dari studi Landi dkk. (2008) mengenai gene expression signature akibat paparan rokok pada kanker paru. Dataset menggunakan platform GPL96 – [HG-U133A] Affymetrix Human Genome U133A Array, yaitu platform microarray satu kanal (single-channel) berbasis probe untuk mengukur ekspresi ±22.000 probe set gen manusia. Sampel terdiri dari jaringan adenokarsinoma paru (Tumor/T) dan jaringan paru non-tumor (Non-Tumor/NT) yang diperoleh dari individu dengan status merokok berbeda (current smokers, former smokers, dan never smokers). Data mentah microarray telah dinormalisasi menggunakan metode Robust Multi-array Average (RMA) sesuai dengan pendekatan pada paper asli. Data diunduh secara langsung menggunakan package GEOquery pada R dan dianalisis menggunakan lingkungan komputasi R serta paket Bioconductor.

2.2 Preprocessing

atriks ekspresi gen diekstraksi dari objek ExpressionSet menggunakan fungsi `exprs()`. Karena platform Affymetrix umumnya telah dinormalisasi dengan metode RMA yang menghasilkan data dalam skala \log_2 , dilakukan pengecekan distribusi nilai ekspresi untuk memastikan apakah transformasi tambahan diperlukan. Jika data telah berada dalam skala \log_2 , maka tidak dilakukan transformasi ulang. Informasi karakteristik sampel, termasuk status merokok dan tipe jaringan (Tumor atau Non-Tumor), diperoleh dari metadata (pData). Variabel kelompok didefinisikan sebagai faktor untuk keperluan analisis statistik. Selanjutnya dibuat design matrix menggunakan fungsi `model.matrix()` sebagai dasar pembangunan model linear dalam analisis diferensial ekspresi.

2.3 Analisis Differential Expression

Analisis diferensial ekspresi gen dilakukan menggunakan package limma (Linear Models for Microarray Data). Langkah-langkah analisis meliputi:

1. Pembangunan model linear untuk setiap gen menggunakan fungsi `lmFit()`.
2. Penentuan kontras antar kelompok (misalnya current vs never smokers) menggunakan fungsi `makeContrasts()`.
3. Penerapan kontras ke model menggunakan `contrasts.fit()`.
4. Moderasi estimasi varians menggunakan metode empirical Bayes melalui fungsi `eBayes()`.

Gen yang mengalami perubahan ekspresi signifikan diidentifikasi menggunakan fungsi `topTable()` dengan kriteria:

- Adjusted p-value (FDR, metode Benjamini-Hochberg) $< 0,01$
- $|\log_2 \text{Fold Change}| > 1$

Gen dengan nilai logFC positif dikategorikan sebagai upregulated, sedangkan gen dengan logFC negatif dikategorikan sebagai downregulated pada kelompok perokok dibandingkan non-perokok.

2.4 Visualisasi

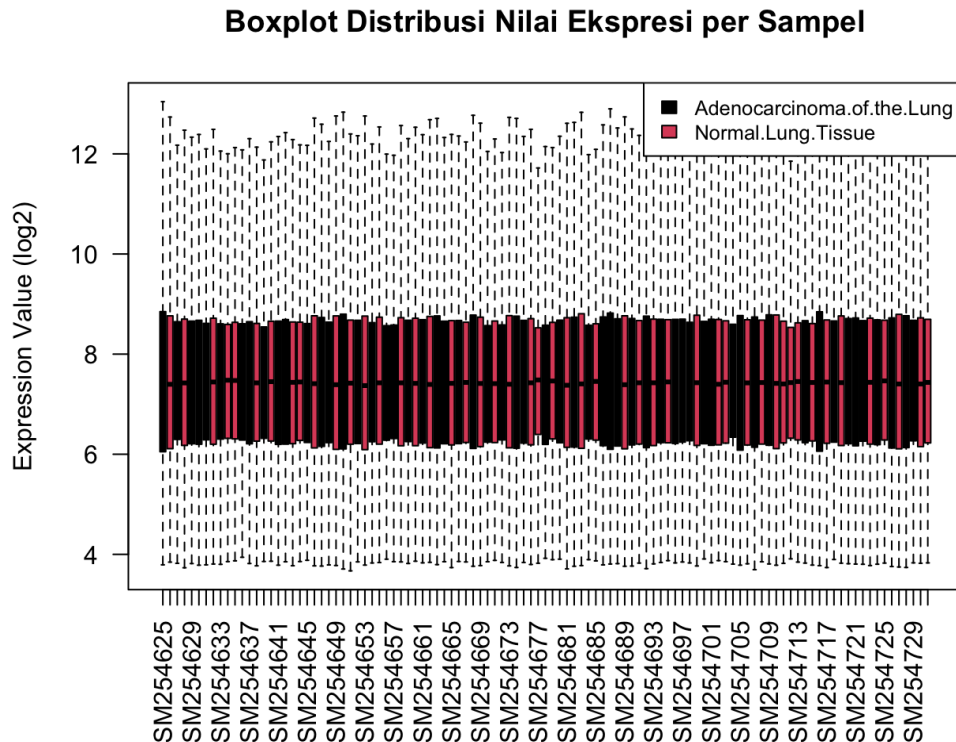
Visualisasi dilakukan untuk mengevaluasi kualitas data serta menginterpretasikan hasil analisis diferensial.

- Boxplot digunakan untuk memeriksa distribusi nilai ekspresi antar sampel dan mengevaluasi keseragaman data.
- Volcano plot digunakan untuk menampilkan distribusi gen berdasarkan nilai log₂ Fold Change dan tingkat signifikansi statistik ($-\log_{10}$ adjusted p-value).
- Heatmap dibuat untuk 50 gen dengan nilai adjusted p-value paling signifikan guna melihat pola ekspresi dan pengelompokan sampel berdasarkan gen diferensial.

2.5 Analisis Enrichment

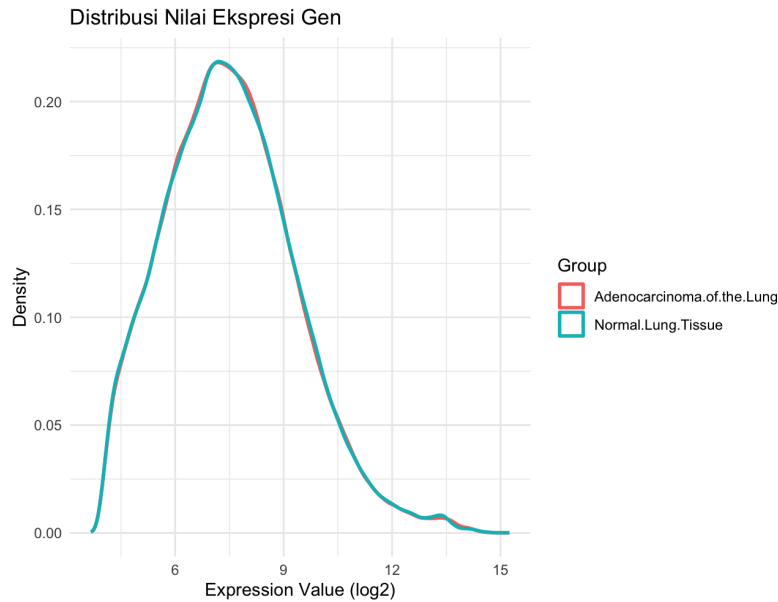
Gen yang teridentifikasi signifikan selanjutnya dianalisis menggunakan analisis enrichment untuk mengidentifikasi jalur biologis yang terlibat. Analisis dilakukan pada kategori **Gene Ontology (Biological Process)** dan **KEGG pathway** menggunakan package seperti *clusterProfiler* atau *enrichR*. Hasil enrichment divisualisasikan dalam bentuk barplot atau dotplot untuk mempermudah interpretasi.

HASIL DAN INTERPRETASI



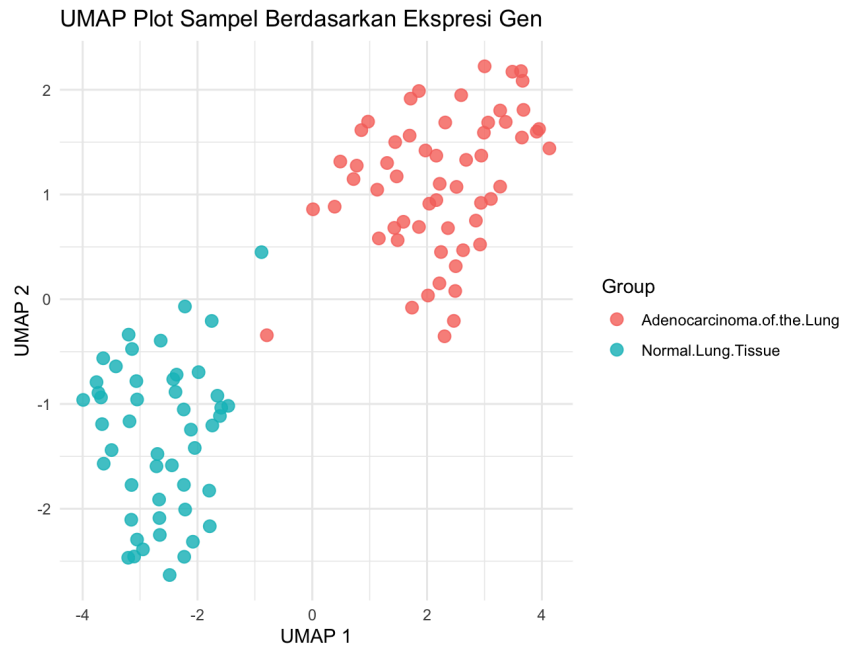
Gambar 1. Grafik Boxplot Ekspresi tiap Sampel

Visualisasi distribusi ekspresi gen melalui grafik boxplot menunjukkan bahwa proses normalisasi data telah berhasil dilakukan dengan sangat baik, di mana nilai median dan rentang interkuartil di seluruh sampel, baik dari kelompok paru-paru normal maupun adenokarsinoma terlihat sejajar dan konsisten. Keseragaman distribusi ini mengonfirmasi bahwa tidak terdapat variasi teknis yang ekstrem atau bias sistematis antar sampel (seperti efek batch), sehingga data memiliki varians yang stabil dan berada dalam rentang skala yang setara untuk dibandingkan secara statistik. Dengan kondisi data yang sudah "rapi" dan tidak menunjukkan adanya *outlier* yang melenceng secara signifikan, hasil analisis ekspresi gen turunan di tahap berikutnya, termasuk identifikasi gen yang berbeda bermakna serta analisis jalur fungsionalnya, memiliki tingkat validitas dan reliabilitas yang tinggi untuk diinterpretasikan secara biologis.



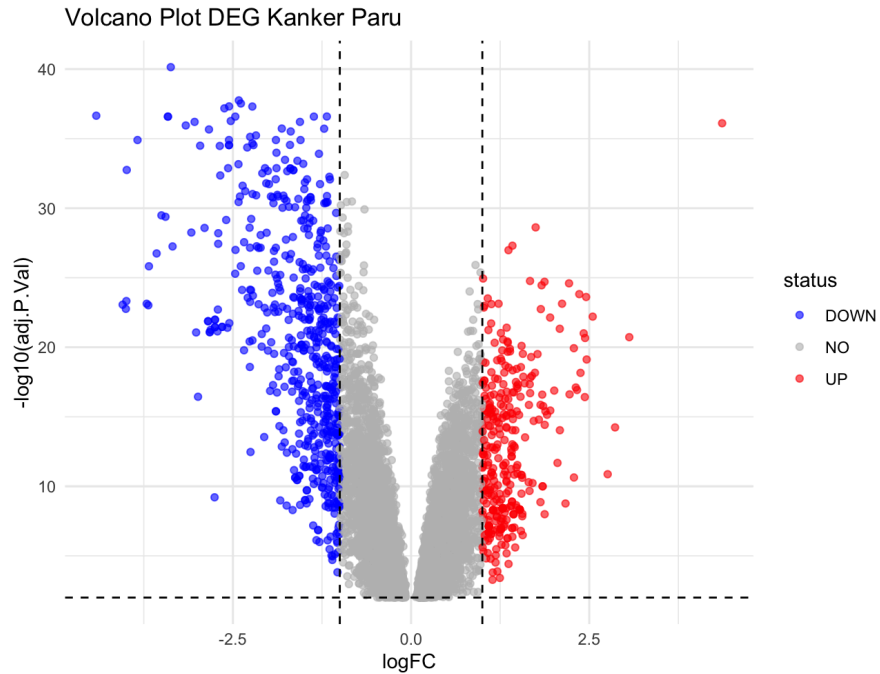
Gambar 2. Grafik Distribusi Ekspresi Gen (Density Plot)

Visualisasi density plot tersebut berfungsi untuk menunjukkan pola sebaran frekuensi nilai ekspresi gen di seluruh sampel, di mana bentuk kurva yang saling berhimpit sempurna mengonfirmasi bahwa data telah terdistribusi secara identik setelah proses normalisasi. Puncak kurva yang tunggal dan simetris menandakan bahwa sebagian besar gen memiliki tingkat ekspresi yang terpusat pada rentang nilai yang sama, sehingga tidak ditemukan adanya anomali distribusi atau pergeseran populasi data yang ekstrem antar sampel. Dalam konteks analisis data besar seperti 9.500 gen ini, keselarasan bentuk "gunung" pada grafik tersebut sangat krusial karena membuktikan bahwa setiap sampel memiliki karakteristik statistik yang setara. Dengan demikian, perbandingan ekspresi gen yang dilakukan di tahap berikutnya tidak akan terbiaskan oleh perbedaan sebaran data mentah, melainkan murni mencerminkan perbedaan biologis antara jaringan normal dan kanker.



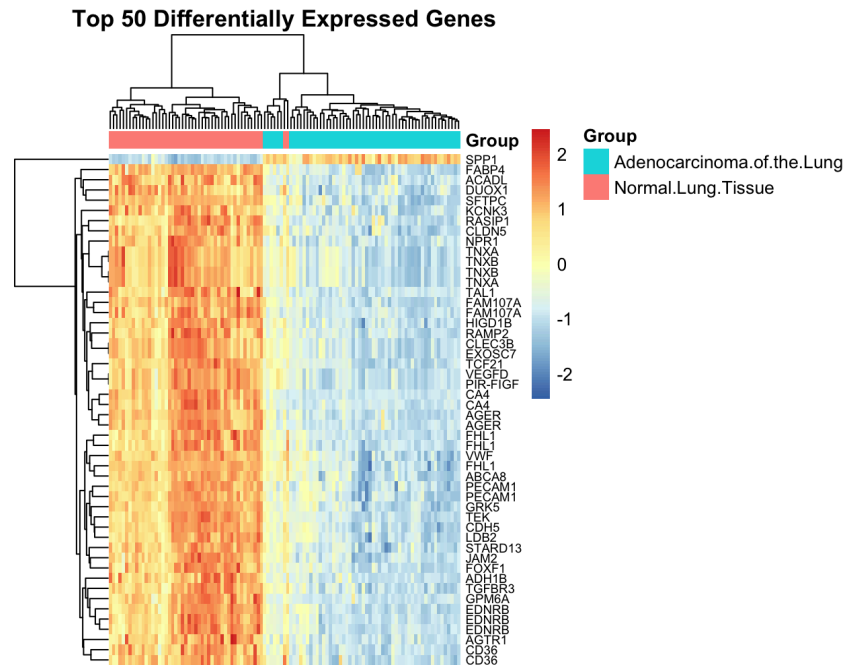
Gambar 3. Grafik UMAP Ekspresi Gen

Visualisasi UMAP (Uniform Manifold Approximation and Projection) ini berfungsi untuk menunjukkan pengelompokan seluruh sampel berdasarkan kemiripan profil ekspresi gennya secara keseluruhan dalam ruang dua dimensi. Terlihat jelas adanya pemisahan yang sangat tegas antara kelompok Adenocarcinoma of the Lung (titik merah) dan kelompok Normal Lung Tissue (titik biru), di mana masing-masing kelompok membentuk kluster yang terpisah jauh tanpa ada tumpang tindih yang berarti. Pemisahan yang kontras ini membuktikan secara visual bahwa terdapat perbedaan tanda genetik (*molecular signature*) yang sangat kuat dan konsisten antara jaringan kanker dan jaringan normal. Hal ini memberikan validasi tambahan bahwa data ekspresi gen yang kamu miliki sangat berkualitas karena mampu membedakan kedua kondisi biologis tersebut secara akurat, sehingga analisis lanjut seperti identifikasi biomarker atau jalur jalur persinyalan yang terganggu menjadi sangat relevan dan memiliki dasar yang kuat.



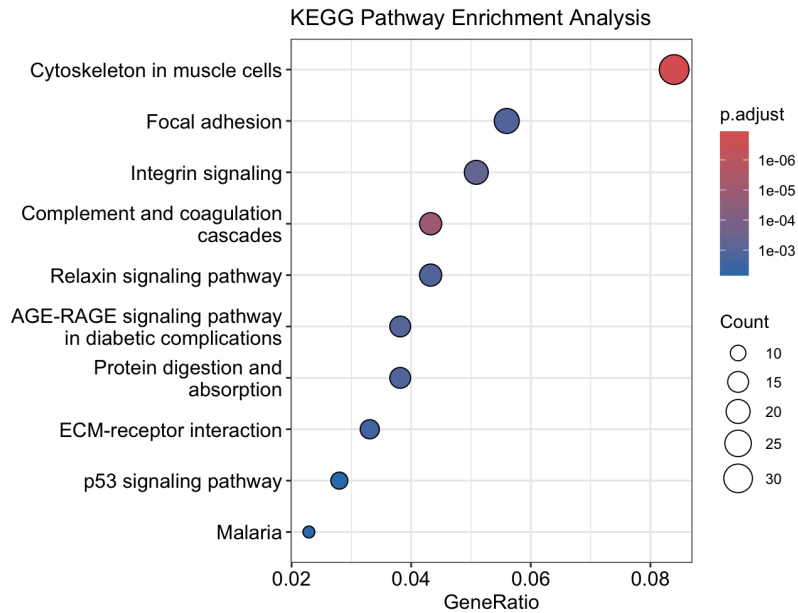
Gambar 4. Grafik Volcano Ekspresi Gen

Visualisasi Volcano Plot ini merangkum hasil identifikasi gen yang terekspresi secara diferensial (DEG) dengan membandingkan tingkat signifikansi statistik terhadap besarnya perubahan ekspresi gen antara kelompok kanker dan normal. Setiap titik pada grafik mewakili satu gen, di mana titik-titik berwarna biru (Downregulated) dan merah (Upregulated) menunjukkan gen-gen yang memiliki nilai signifikansi tinggi (di atas garis horizontal putus-putus) serta perubahan lipatan (*fold change*) yang besar (di luar garis vertikal putus-putus). Sebaran titik yang lebih condong ke arah kiri (biru) menunjukkan bahwa dalam dataset ini terdapat lebih banyak gen yang mengalami penurunan ekspresi pada kondisi Adenokarsinoma Paru dibandingkan gen yang mengalami peningkatan. Gen-gen yang berada di posisi paling atas dan paling ujung (kiri atau kanan) merupakan kandidat biomarker paling potensial karena menunjukkan perbedaan yang sangat kontras dan konsisten secara statistik, yang kemudian menjadi dasar bagi analisis fungsional jalur KEGG dan GO yang telah dilakukan sebelumnya.



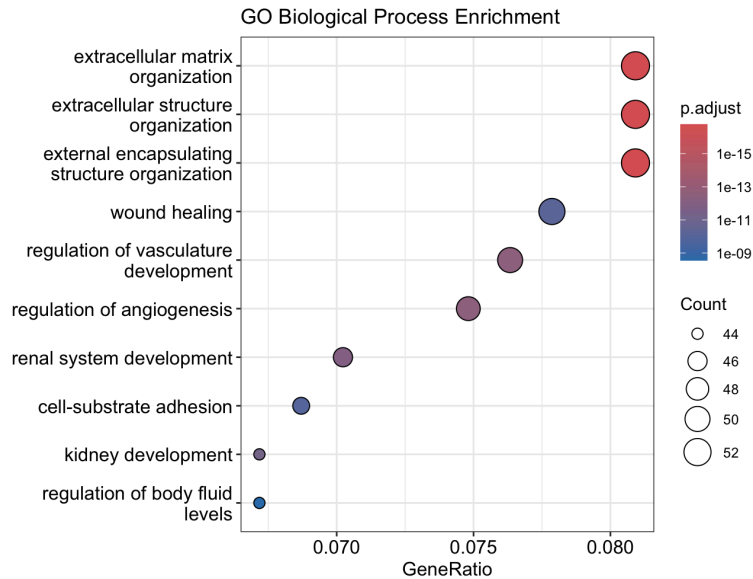
Gambar 5. Grafik Heatmap Ekspresi Gen

Visualisasi Heatmap Top 50 Differentially Expressed Genes (DEGs) ini memberikan gambaran yang sangat jelas mengenai perbedaan profil molekuler antara jaringan Adenokarsinoma Paru (berwarna biru pada anotasi atas) dan Jaringan Paru Normal (berwarna merah pada anotasi atas). Warna pada heatmap mewakili tingkat ekspresi gen, di mana warna merah menunjukkan ekspresi tinggi dan warna biru menunjukkan ekspresi rendah. Terlihat kontras yang sangat tajam secara horizontal: gen-gen yang terekspresi sangat kuat di jaringan normal justru meredup di jaringan kanker, dan sebaliknya. Pola pengelompokan (*dendrogram*) di bagian atas juga menunjukkan bahwa sampel-sampel dalam kelompok yang sama berkumpul menjadi satu kluster besar, yang membuktikan bahwa perbedaan ekspresi pada 50 gen teratas ini sangat konsisten dan mampu menjadi pembeda (biomarker) yang akurat. Secara biologis, daftar gen di sisi kanan, seperti *AGER*, *CA4*, dan *VWF*, yang menunjukkan penurunan drastis pada kanker, sering kali berkaitan dengan hilangnya fungsi normal jaringan paru, sementara gen yang meningkat berkaitan dengan agresivitas tumor. Secara keseluruhan, heatmap ini memvalidasi hasil *Volcano Plot* sebelumnya dengan menunjukkan bahwa perbedaan signifikan tersebut bukan hanya angka statistik, melainkan pola nyata yang memisahkan kondisi sehat dari kondisi kanker secara visual.



Gambar 6. Grafik Enrichment KEGG

Visualisasi KEGG Pathway Enrichment Analysis dalam bentuk dotplot ini menunjukkan jalur-jalur biologis utama yang paling terpengaruh oleh perubahan ekspresi gen pada kasus adenokarsinoma paru. Jalur Cytoskeleton in muscle cells muncul sebagai jalur yang paling signifikan, ditandai dengan titik merah besar yang memiliki nilai GeneRatio tertinggi (di atas 0.08) dan nilai p.adjust yang sangat rendah (mendekati 1×10^{-6}). Selain itu, munculnya jalur-jalur seperti Focal adhesion, Integrin signaling, dan ECM-receptor interaction (ditandai dengan titik berwarna biru hingga keunguan) menunjukkan bahwa proses keganasan pada sel kanker paru sangat berkaitan erat dengan perubahan struktur mekanik sel dan interaksinya dengan lingkungan sekitar (mikrolingkungan tumor). Jalur persinyalan penting seperti p53 signaling pathway juga terdeteksi, yang mengonfirmasi adanya gangguan pada mekanisme kontrol siklus sel dan perbaikan DNA yang umum ditemukan pada pertumbuhan kanker. Secara keseluruhan, grafik ini memberikan bukti kuat bahwa modulasi pada arsitektur seluler dan jalur persinyalan pertumbuhan adalah pendorong utama perbedaan molekuler antara jaringan sehat dan jaringan kanker pada dataset Anda.



Gambar 7. Grafik Enrichment GO

Visualisasi GO Biological Process Enrichment dalam bentuk dotplot ini memberikan pandangan yang lebih mendalam mengenai proses biologis spesifik yang teraktivasi pada sampel adenokarsinoma paru. Jalur yang berkaitan dengan modifikasi lingkungan seluler, seperti extracellular matrix organization dan extracellular structure organization, menempati posisi teratas dengan tingkat signifikansi statistik yang sangat ekstrem (ditandai dengan warna merah menyala dengan nilai p.adjust mencapai 10⁻¹⁵). Tingginya nilai GeneRatio dan Count (lebih dari 52 gen) pada jalur-jalur ini menunjukkan bahwa perombakan struktur luar sel adalah fenomena molekuler yang paling dominan dalam dataset Anda.

Selain itu, grafik ini menyoroti adanya aktivitas signifikan pada proses wound healing (penyembuhan luka), regulation of vasculature development, dan angiogenesis. Munculnya jalur-jalur ini mengindikasikan bahwa sel kanker tidak hanya mengubah struktur fisiknya, tetapi juga memicu pembentukan pembuluh darah baru untuk mendukung suplai nutrisi dan pertumbuhan tumor. Secara keseluruhan, dotplot GO ini memperkuat temuan KEGG sebelumnya, menegaskan bahwa perkembangan kanker pada pasien ini sangat didorong oleh mekanisme perombakan matriks ekstraseluler dan manipulasi sistem vaskular tubuh untuk memfasilitasi invasi sel kanker.

KESIMPULAN

Berdasarkan hasil analisis bioinformatika terhadap dataset ekspresi gen adenokarsinoma paru, dapat disimpulkan bahwa transisi dari jaringan paru normal menjadi ganas disertai dengan perubahan molekuler yang sistematis dan signifikan. Analisis statistik berhasil mengidentifikasi profil *Differentially Expressed Genes* (DEGs) yang secara konsisten mampu memisahkan kelompok kanker dari jaringan normal, sebagaimana divalidasi melalui visualisasi UMAP dan *Heatmap* Top 50 DEGs. Pola ekspresi yang kontras pada gen-gen utama seperti *AGER*, *CA4*, dan *VWF* menunjukkan hilangnya identitas fungsional jaringan paru normal pada sel kanker, sementara distribusi pada *Volcano Plot* mengonfirmasi dominasi gen yang mengalami penurunan ekspresi (*downregulated*) pada kondisi tumor.

Lebih lanjut, analisis pengayaan fungsional melalui KEGG dan *Gene Ontology* (GO) mengungkap bahwa mekanisme patogenesis utama pada adenokarsinoma paru ini didorong oleh reorganisasi arsitektur seluler dan perombakan mikrolingkungan tumor. Aktivitas dominan pada jalur *extracellular matrix organization*, *focal adhesion*, dan *angiogenesis* menunjukkan bahwa sel kanker secara aktif memodifikasi struktur sekitarnya dan memicu pembentukan pembuluh darah baru untuk mendukung invasi serta metastasis. Integrasi hasil ini menegaskan bahwa tanda genetik yang teridentifikasi bukan sekadar fluktuasi statistik, melainkan mencerminkan strategi biologis sel kanker dalam memanipulasi mekanisme penyembuhan luka dan struktur mekanik jaringan untuk memfasilitasi progresivitas penyakit.

REFERENSI

Landi, M. T., Dracheva, T., Rotunno, M., Figueroa, J. D., Liu, H., Dasgupta, A., Mann, F. E., Fukuoka, J., Hames, M., Bergen, A. W., Murphy, S. E., Yang, P., Pesatori, A. C., Consonni, D., Bertazzi, P. A., Wacholder, S., Shih, J. H., Caporaso, N. E., & Jen, J. (2008). Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS ONE*, 3(2), e1651. <https://doi.org/10.1371/journal.pone.0001651>